Peatükk 9

Osamudeli hindamisest

Vaatame lineaarset mudelit,

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

mille parameetervektor on jagatud kaheks osaks $\boldsymbol{\beta}^T = \left(\boldsymbol{\beta}_1^T | \boldsymbol{\beta}_2^T\right)$. Samuti saame jagada ka mudelimaatriksi kaheks parameetervektori tükile vastavaks osaks, $\mathbf{X} = (\mathbf{X}_1|\mathbf{X}_2)$, ning soovi korral võime algse mudeli kirja panna kujul

$$\mathbf{y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon},$$

Soovime leida, millised näevad välja hinnangud $\boldsymbol{\beta}_1$-le ja $\boldsymbol{\beta}_2$-le. Antud juhul teeme arutelu läbi eeldusel, et parameetervektor on hinnatav (ehk maatriks $\mathbf{X}^T\mathbf{X}$ on pööratav).

Esmalt juhime tähelepanu mõnele maatriksalgebra tulemusele. Blokkmaatriksi pöördmaatriksit on võimalik leida järgmise valemi abil:

$$\begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{pmatrix}^{-1} = \begin{pmatrix} \left(\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C}\right)^{-1} & -\left(\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C}\right)^{-1}\mathbf{B}\mathbf{D}^{-1} \\ -\left(\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B}\right)^{-1}\mathbf{C}\mathbf{A}^{-1} & \left(\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B}\right)^{-1} \end{pmatrix}.$$

Paneme kirja, milline näeb välja parameetervektori $\boldsymbol{\beta}$ hinnang:

$$\begin{aligned} \hat{\boldsymbol{\beta}} &= \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{y} \\ &= \begin{pmatrix} \mathbf{X}_1^T\mathbf{X}_1 & \mathbf{X}_1^T\mathbf{X}_2 \\ \mathbf{X}_2^T\mathbf{X}_1 & \mathbf{X}_2^T\mathbf{X}_2 \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{X}_1^T \\ \mathbf{X}_2^T \end{pmatrix} \mathbf{y} \end{aligned}$$

Leiame, milline näeb välja pöördmaatriks, kasutades ülaltoodud valemit blokkmaatriksi pöördmaatriksi leidmiseks:

$$\begin{aligned}
\left(\mathbf{A}-\mathbf{B}\mathbf{D}^{-1}\mathbf{C}\right)^{-1} &= \left(\mathbf{X}_1^T\mathbf{X}_1-\mathbf{X}_1^T\mathbf{X}_2\left(\mathbf{X}_2^T\mathbf{X}_2\right)^{-1}\mathbf{X}_2^T\mathbf{X}_1\right)^{-1}\\
&= \left(\mathbf{X}_1^T\left(\mathbf{I}-\mathbf{P}_{\mathbf{X}_2}\right)\mathbf{X}_1\right)^{-1}
\end{aligned}$$

$$-\left(\mathbf{A}-\mathbf{B}\mathbf{D}^{-1}\mathbf{C}\right)^{-1}\mathbf{B}\mathbf{D}^{-1} = \left(\mathbf{X}_1^T\left(\mathbf{I}-\mathbf{P}_{\mathbf{X}_2}\right)\mathbf{X}_1\right)^{-1}\mathbf{X}_1^T\mathbf{X}_2\left(\mathbf{X}_2^T\mathbf{X}_2\right)^{-1}$$

$$\begin{aligned}
\left(\mathbf{D}-\mathbf{C}\mathbf{A}^{-1}\mathbf{B}\right)^{-1} &= \left(\mathbf{X}_2^T\mathbf{X}_2-\mathbf{X}_2^T\mathbf{X}_1\left(\mathbf{X}_1^T\mathbf{X}_1\right)^{-1}\mathbf{X}_1^T\mathbf{X}_2\right)^{-1}\\
&= \left(\mathbf{X}_2^T\left(\mathbf{I}-\mathbf{P}_{\mathbf{X}_1}\right)\mathbf{X}_2\right)^{-1}
\end{aligned}$$

$$-\left(\mathbf{D}-\mathbf{C}\mathbf{A}^{-1}\mathbf{B}\right)^{-1}\mathbf{C}\mathbf{A}^{-1} = \left(\mathbf{X}_2^T\left(\mathbf{I}-\mathbf{P}_{\mathbf{X}_1}\right)\mathbf{X}_2\right)^{-1}\mathbf{X}_2^T\mathbf{X}_1\left(\mathbf{X}_1^T\mathbf{X}_1\right)^{-1}$$

ja lõplikuks hinnanguks saame

$$\begin{pmatrix}\hat{\boldsymbol\beta}_1\\ \hat{\boldsymbol\beta}_2\end{pmatrix} = \begin{pmatrix}\left(\mathbf{X}_1^T\left(\mathbf{I}-\mathbf{P}_{\mathbf{X}_2}\right)\mathbf{X}_1\right)^{-1}\mathbf{X}_1^T\left(\mathbf{I}-\mathbf{P}_{\mathbf{X}_2}\right)\mathbf{y}\\ \left(\mathbf{X}_2^T\left(\mathbf{I}-\mathbf{P}_{\mathbf{X}_1}\right)\mathbf{X}_2\right)^{-1}\mathbf{X}_2^T\left(\mathbf{I}-\mathbf{P}_{\mathbf{X}_1}\right)\mathbf{y}\end{pmatrix} \tag{9.1}$$

Interpretatsioon: parameetervektori esimese poole hinnangu saamiseks peaksime eemaldama muutujate $\mathbf{X}_2$ mõju muutjatest $\mathbf{X}_1$ ja seejärel hindama saadud jääkide mõju sõltuvale tunnusele.

Alljärgnevalt toome ära ühe arutluskäigu, mille tulemuste ekslik tõlgendamine viib sageli tõsiste eksimusteni reaalsete andmete analüüsimisel. Uurime nimelt, kuna saame mudelimaatriksit $\mathbf{X}_1$ kasutades head (nihketa) hinnangud parameetritele $\boldsymbol\beta_1$ (ja kuna peame tingimata kasutama ka teist poolt mudelimaatriksist ($\mathbf{X}_2$-te). Täpsemalt: olgu $\mathbf{y} = \mathbf{X}_1\boldsymbol\beta_1 + \mathbf{X}_2\boldsymbol\beta_2 + \varepsilon$. Kui hindame parameetervektori $\boldsymbol\beta_1$ kasutades lihtsamat mudelit ($\mathbf{y} = \mathbf{X}_1\boldsymbol\beta_1 + \varepsilon$), siis millal saame nihketa hinnangu meid huvitavataele parameetritele (vaatamata sellele, et kasutasime vale mudelit)? Vaatame seda küsimust veidi lähemalt:

$$\begin{aligned}
\mathrm{E}\left((\mathbf{X}_1^T\mathbf{X}_1)^{-1}\mathbf{X}_1\mathbf{y}\right) &= \mathrm{E}\left((\mathbf{X}_1^T\mathbf{X}_1)^{-1}\mathbf{X}_1(\mathbf{X}_1\boldsymbol\beta_1+\mathbf{X}_2\boldsymbol\beta_2+\varepsilon)\right)\\
&= \boldsymbol\beta_1 + (\mathbf{X}_1^T\mathbf{X}_1)^{-1}\mathbf{X}_1^T\mathbf{X}_2\boldsymbol\beta_2.
\end{aligned}$$

Saadud avaldis võrdub $\boldsymbol{\beta}_1$-ga siis, kui $\mathbf{X}_2 \perp \mathbf{X}_1$ või kui $\boldsymbol{\beta}_2 = 0$ (kui mudelist väljajäänud tunnused on kas ortogonaalsed — või sõltumatud — mudelisse sattunud tunnustest või kui mudelist väljajäänud tunnused tegelikult uuritava tunnuse keskväärtust ei mõjuta).

Sageli üritatakse saadud tulemust aga valesti ära kasutada. Nimelt üritatakse defineerida uut uuritavat tunnust, kust on eemaldatud nn teise blokki jäävate tunnuste mõju, $\mathbf{y}^* = \mathbf{y} - \mathbf{X}_2\boldsymbol{\beta}_2$. Selliselt defineeritud tunnus ei sõltu enam mudelimaatriksisse $\mathbf{X}_2$ jäänud tunnuste mõjust ja seega võiksime uurida tunnuste $\mathbf{y}^*$ ja $\mathbf{X}_1$ vahelist seost ilma maatriksit $\mathbf{X}_2$ kasutamata. Reaalses elus pole aga $\boldsymbol{\beta}_2$ teada. Praktikas tehaksegi nüüd $\boldsymbol{\beta}_2$ hindamisel viga — hinnatakse see vaid vektorit $\mathbf{y}$ ja maatriksit $\mathbf{X}_2$ kasutades ja saadakse seega nihkega hinnang $\boldsymbol{\beta}_2$-le, mis viib kogu edasise analüüsi omadega metsa.

Saadud tulemusest 9.1 võib olla kasu ka mõistmaks, miks ühe või teise parameetri hindamistäpsus on madal või aru saamaks mida tuleks teha saamaks võimalikult täpset parameetri hinnangut.

Vaatame juhtu, kui $\boldsymbol{\beta}_1$ on vektor pikkusega 1 (tegemist on üheainsa parameetriga).

$$\begin{aligned} \mathrm{D}\left(\hat{\beta}_1\right) &= \left(\mathbf{X}_1^T\left(\mathbf{I} - \mathbf{P}_{\mathbf{X}_2}\right)\mathbf{X}_1\right)^{-1}\sigma^2 \\ &= \frac{\sigma^2}{\mathbf{X}_1^T\left(\mathbf{I} - \mathbf{P}_{\mathbf{X}_2}\right)\mathbf{X}_1} \end{aligned}$$

Paneme tähele, et $SSE_{\mathbf{X}_1\sim\mathbf{X}_2} := \mathbf{X}_1^T\left(\mathbf{I} - \mathbf{P}_{\mathbf{X}_2}\right)\mathbf{X}_1$ on jääkide ruutude summa mudelis, kus prognoositakse tunnuse $\mathbf{X}_1$ väärtust kasutades mudeli maatriksit $\mathbf{X}_2$. Samuti teame, et

$$\begin{aligned} R^2_{\mathbf{X}_1\sim\mathbf{X}_2} &= 1 - \frac{SSE_{\mathbf{X}_1\sim\mathbf{X}_2}}{SSE_{\mathbf{X}_1\sim 1}} \\ SSE_{\mathbf{X}_1\sim\mathbf{X}_2} &= SSE_{\mathbf{X}_1\sim 1}(1 - R^2_{\mathbf{X}_1\sim\mathbf{X}_2}). \end{aligned}$$

Seega võime parameetri β_1 hinnagu dispersiooni kirja panna järgmisel kujul:

$$\begin{aligned} \mathrm{D}\left(\hat{\beta}_1\right) &= \frac{\sigma^2}{SSE_{\mathbf{X}_1\sim\mathbf{X}_2}} \\ &= \frac{\sigma^2}{SSE_{\mathbf{X}_1\sim 1}} \cdot \frac{1}{1 - R^2_{\mathbf{X}_1\sim\mathbf{X}_2}} \\ &= \frac{1}{n-1} \cdot \frac{\sigma^2}{\hat{\sigma}^2_{x_1}} \cdot \frac{1}{1 - R^2_{\mathbf{X}_1\sim\mathbf{X}_2}}. \end{aligned}$$

Suurust $\frac{1}{1-R^2_{\mathbf{x}_1 \sim \mathbf{x}_2}}$ tuntakse (tunnuse X_1) hinnangu dispersiooni puhitusteguri nime all (*Variance Inflation Factor, VIF*) ja ta iseloomustab seda, kuivõrd kaotame hinnangu täpsuses seetõttu, et kasutame mitteortogonaalseid (omavahel korreleeritud) sõltuvaid tunnuseid mudeli hindamisel.

Ülesanne

Hinnatakse kolme tunnust sisaldav mudel. Kõik tunnused osutuvad statistiliselt olulisteks, aga paari tunnuse dispersiooni puhitustegurid on ääretult suured:

```
tunnus    VIF
X1        1.0
X2       17.0
X3       17.0
```

Kas tunnuse X1 mõjule leitud hinnang võiks täpsemaks minna (kas hinnangu standardviga väheneks) kui mudelist eemaldata ülimalt suure VIF väärtusega tunnus X3? Põhjenda oma arvamust!